

WHAT IS CLAIMED IS:

1. A method of energy management in a computer system having a plurality of computation nodes comprising the steps of:

assigning a first computation node to an Operational node set as an Operational node, wherein said first computation node is a fully active node;

assigning a second computation node to a Standby node set as a Standby node, wherein said second computation node has its processor(s) and memory in a minimum power consumption state corresponding to maintaining essential data; and

assigning remaining of said plurality of computation nodes excluding said first and second nodes to a Hibernating node set as hibernating nodes, wherein hibernating nodes are maintained in a powered down state.

2. The method of claim 1 further comprising the steps of:

setting a lower computational workload limit (WL2) and an upper computational workload limit (WL1) for said first computation node; and

comparing an actual average workload (WL) of said first computation node to said WL2 and said WL1.

3. The method of claim 2 further comprising the steps of:

redistributing the workload of said first computation node to a third computation node in said Operational node set when said WL of said first computation node is less than WL2; and

moving said first computation node to said Hibernating node set.

4. The method of claim 2 further comprising the step of:

moving workload from said first computation node to a third computation node

3 when said WL of said first computation node is greater than WL1 such that said WL of
4 said first computation node and a WL of said third computation node both are less than
5 WL1.

1 5. The method of claim 2 further comprising the steps of:

2 moving a fifth computation node from said Hibernating node set to said Standby
3 node set in response to a determination that said WL of said first node is greater than
4 WL1;

5 moving a sixth computation node from said Standby node set to said Operational
6 node set response to said determination that said WL of said first node is greater than
7 WL1; and

8 redistributing workload from said first computation node to said sixth
9 computation node such that said WL of said first computation node and a WL of said
10 sixth computation node are both less than WL1.

1 6. The method of claim 1, wherein said computer system is a massively parallel
2 processors system (MPP).

1 7. The method of claim 6, wherein said computation node comprises a single
2 processor.

1 8. The method of claim 1, wherein said computer system is a symmetrical
2 multiprocessor system (SMP).

1 9. The method of claim 8, wherein said computation node comprises multiple
2 processors coupled to a shared memory unit.

- 1 10. The method of claim 1, wherein said first computation node executes a process
2 to minimize energy consumption by a combination of voltage and frequency scaling,
3 wherein said minimized energy consumption enables a required performance of said first
4 computation node.

1 11. A computer program product for, said computer program product embodied in
2 a machine readable medium for energy management in a computer system having a
3 plurality of computation nodes, including programming for a processor, said computer
4 program comprising a program of instructions for performing the program steps of:

5 assigning a first computation node to an Operational node set as an Operational
6 node, wherein said first computation node is a fully active node;

7 assigning a second computation node to a Standby node set as a Standby node,
8 wherein said second computation node has its processor(s) and memory in a minimum
9 power consumption state corresponding to maintaining essential data; and

10 assigning remaining of said plurality of computation nodes excluding said first
11 and second nodes to a Hibernating node set as hibernating nodes, wherein hibernating
12 nodes are maintained in a powered down state.

1 12. The computer program product of claim 11 further comprising the program steps
2 of:

3 setting a lower computational workload limit (WL2) and a upper computational
4 workload limit (WL1) for said first computation node; and

5 comparing an actual average workload (WL) of said first computation node to
6 said WL2 and said WL1.

1 13. The computer program product of claim 12 further comprising the program steps
2 of:

3 redistributing the workload of said first computation node to a third computation
4 node in said Operational node set when said WL of said first computation node is less
5 than WL2; and

6 moving said first computation node to said Hibernating node set.

7 14. The computer program product of claim 12 further comprising the program step
8 of:

9 moving workload from said first computation node to a third computation node
10 when said WL of said first computation node is greater than WL1 such that said WL of
11 said first computation node and a WL of said third computation node both are less than
12 WL1.

1 15. The computer program product of claim 12 further comprising the program steps
2 of:

3 moving a fifth computation node from said Hibernating node set to said Standby
4 node set in response to a determination that said WL of said first node is greater than
5 WL1;

6 moving a sixth computation node from said Standby node set to said Operational
7 node set response to said determination that said WL of said first node is greater than
8 WL1; and

9 redistributing workload from said first computation node to said sixth
10 computation node such that said WL of said first computation node and a WL of said
11 sixth computation node are both less than WL1.

1 16. The computer program product of claim 11, wherein said computer system is a
2 massively parallel processors system (MPP).

1 17. The computer program product of claim 16, wherein said computation node
2 comprises a single processor.

1 18. The computer program product of claim 11, wherein said computer system is a
2 symmetrical multiprocessor system (SMP).

3 19. The computer program product of claim 18, wherein said computation node
4 comprises multiple processors coupled to a shared memory unit.

1 20. The computer program product of claim 11, wherein said first computation node
2 executes a process to minimize energy consumption by a combination of voltage and
3 frequency scaling, wherein said minimized energy consumption enables a required
4 performance of said first computation node.

5 21. A system for energy management in a computer system having a plurality of
6 computation nodes comprising:

7 circuitry for assigning a first computation node to an Operational node set as an
8 Operational node, wherein said first computation node is a fully active node;

9 circuitry for assigning a second computation node to a Standby node set as a
10 Standby node, wherein said second computation node has its processor(s) and memory
11 in a minimum power consumption state corresponding to maintaining essential data; and

12 circuitry for assigning remaining of said plurality of computation nodes excluding
13 said first and second nodes to a Hibernating node set as hibernating nodes, wherein
14 hibernating nodes are maintained in a powered down state.

1 22. The system of claim 21 further comprising:

2 circuitry for setting a lower computational workload limit (WL2) and a upper
3 computational workload limit (WL1) for said first computation node; and

4 circuitry for comparing an actual average workload (WL) of said first
5 computation node to said WL2 and said WL1.

1 23. The system of claim 22 further comprising:

2 circuitry for redistributing the workload of said first computation node to a third
3 computation node in said Operational node set when said WL of said first computation
4 node is less than WL2; and

5 circuitry for moving said first computation node to said Hibernating node set.

1 24. The system of claim 22 further comprising:

2 circuitry for moving workload from said first computation node to a third
3 computation node when said WL of said first computation node is greater than WL1 such

4 that said WL of said first computation node and a WL of said third computation node
5 both are less than WL1.

1 25. The system of claim 22 further comprising:

2 circuitry for moving a fifth computation node from said Hibernating node set to
3 said Standby node set in response to a determination that said WL of said first node is
4 greater than WL1;

5 circuitry for moving a sixth computation node from said Standby node set to said
6 Operational node set response to said determination that said WL of said first node is
7 greater than WL1; and

8 circuitry for redistributing workload from said first computation node to said
9 sixth computation node such that said WL of said first computation node and a WL of
10 said sixth computation node are both less than WL1.

1 26. The system of claim 21, wherein said computer system is a massively parallel
2 processors system (MPP).

1 27. The system of claim 26, wherein said computation node comprises a single
2 processor.

3 28. The system of claim 21, wherein said computer system is a symmetrical
4 multiprocessor system (SMP).

1 29. The system of claim 28, wherein said computation node comprises multiple
2 processors coupled to a shared memory unit.

1 30. The system of claim 21, wherein said first computation node executes a process
2 to minimize energy consumption by a combination of voltage and frequency scaling,
3 wherein said minimized energy consumption enables a required performance of said first
4 computation node.